Moment Approximation and Estimation of Diffusion Models of Asset Prices

by

Hua He

March 1990

Research Program in Finance
Working Papers are preliminary in nature: their purpose is to stimulate discussion and comment. Therefore, they should not be cited or quoted in any publication without the permission of the author. Single copies of a paper may be requested from the Institute of Business and Economic Research, 156 Barrows Hall.

Institute of Business and Economic Research
University of California at Berkeley
RESEARCH PROGRAM IN FINANCE AT THE
WALTER A. HAAS SCHOOL OF BUSINESS,
UNIVERSITY OF CALIFORNIA, BERKELEY

The Research Program in Finance in the Walter A. Haas School of Business at the University of California has as its purpose the conduct and encouragement of research in finance, investments, banking, securities markets, and financial institutions. The present reprint and working paper series were established in 1971 in conjunction with a grant from the Dean Witter Foundation.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Carl Shapiro, Director

The Institute of Business and Economic Research is an organized research unit at the University of California, Berkeley, whose mission is to promote research by faculty and graduate students in the fields of business and economics. The Institute carries out its mission by organizing programs and activities that enrich the research environment, administering extramural research awards, publishing working papers, and making direct grants for research.
MOMENT APPROXIMATION AND ESTIMATION OF DIFFUSION MODELS OF ASSET PRICES

Hua He

Haas School of Business
University of California, Berkeley
Berkeley, CA 94720

December 1989

Last revised: March 1990

Finance Working Paper #193

*I thank Andrew Lo, Chi-fu Huang, Neil Pearson, and Robert Merton for helpful conversations and comments. Financial support from the Berkeley Program in Finance is gratefully acknowledged.
Abstract

This paper develops a generalized method of moments (GMM) estimator of the parameters of a time-homogenous diffusion process using discretely sampled data. When the functional forms of moments are unknown, GMM estimation is impossible. The paper introduces a numerical approximation of the conditional moments and proposes that GMM estimation be carried out by replacing the true conditional moments by their numerical approximates. Approximation for conditional moments is obtained through approximating the continuous time diffusion process by a sequence of finely spaced discrete time binomial (or multinomial) processes. Large sample properties are investigated for the GMM estimators obtained by this method. The approach is applied to a mean reverting square root process. Monte-Carlo simulations are performed to examine the finite sample properties of the GMM estimators.

KEYWORDS: Moment approximation, diffusion processes, asset prices, Generalized Methods of Moment estimation, and Monte-Carlo experiments.

Hua He
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720
1 Introduction

One of the most important developments in financial economic theory over the past twenty years has been the study of intertemporal behavior of asset prices using continuous time stochastic processes. In the Intertemporal Capital Asset Pricing Model (ICAPM), developed by Merton (1973) and extended by Breeden (1979) and Cox, Ingersoll and Ross (1985), it is assumed that the equilibrium asset prices and a vector of exogenously specified state variables form a multi-dimensional diffusion process. In a general equilibrium (pure exchange economy) setting, Huang (1987) provides sufficient conditions for the equilibrium prices and state variable processes to form a diffusion process. However, to date relatively little research in economics has been devoted to the econometric estimation problems associated with such diffusion processes.

Standard estimation procedures for diffusion processes often require that the transition density function of the underlying process be analytically tractable so that a maximum likelihood (ML) estimator can be applied to a series of discretely sampled observations, see, for example, Billingsley (1969), Marsh and Rosenfeld (1983), and Lo (1986). However, except for a limited class of diffusion processes, it is difficult to obtain a closed form representation of the transition density of an arbitrary diffusion process. Although the transition density function is determined by the Fokker-Planck partial differential equation (PDE), solving the PDE numerically is clearly a formidable task, especially when the dimension of the vector of underlying prices and state variables is large. Alternatively, one may use a generalized method of moments (GMM) procedure to estimate the parameters of diffusion processes using discretely sampled data. An important feature of GMM estimation is that it allows one to perform a specification test through a test of the over-identifying restrictions. GMM estimation suffers a difficulty similar to the difficulty encountered in maximum likelihood estimation, i.e., it is difficult to find analytic expressions for the moment functions of an arbitrary diffusion process.\(^1\)

In this paper we develop a relatively simple numerical procedure that allows us to approximate the conditional moments of time-homogenous diffusion processes. We propose that GMM estimation be carried out by replacing the true conditional moments by their numerical approximates. The basic idea of our approach is similar to the binomial option pricing approach of Cox, Ross and Rubinstein (1979) who approximate the Black-Scholes option pricing formula with the option prices obtained from discrete time binomial processes. Specifically, we approximate the diffusion

\(^1\) Many statistical studies have focused on estimating diffusion process using continuous records, cf., Basawa and Rao (1980), Brown and Hewitt (1978), Kutoyants (1977), and Tsitsiashvili (1977). These studies allow the unknown parameters to appear only in the drift term of the stochastic differential equation, which rules out the situation we are interested in. Although continuous records are available for asset prices, they are subject to severe measurement errors due to the bid-ask spread.
process between any two consecutive sampling dates by a sequence of finely spaced discrete time binomial or multinomial processes. In the approximating sequence, the conditional moments of the next observation, conditional on the current observation, can be computed easily and they converge to the true conditional moments of the limiting diffusion process. Our approximation procedure is easy to apply even for multi-dimensional diffusion processes. This approach, therefore, allows us to apply GMM estimation to a series of discretely sampled observations without knowing the functional form of the conditional moments.

A closely related work on parametric estimation of diffusion asset prices using GMM is the paper by Duffie and Singleton (1989). Duffie and Singleton develop a simulated moments estimator (SME), which involves approximating the steady state (unconditional) moments through simulation of sample paths of the underlying process. In the simulation of a sample path of a diffusion process, they employ a discretization procedure. As a result, the moment approximation in their model is subject to both discretization errors and simulation errors. To control for the simulation errors, the underlying diffusion process needs to be simulated on an sufficiently long time interval. The length of time interval affects the discretization errors, however. Although on any finite time interval the discretization errors can be bounded, these bounds vary with the length of time interval on which the discretization is performed. Thus, the discretization and simulation errors cannot be controlled simultaneously. This forces Duffie and Singleton to assume a discrete time Markovian structure for the dynamics of asset prices. The procedure of this paper avoids the complication of controlling both discretization and simulation errors, and the moment approximation in our model is subject only to discretization errors.\footnote{Bossaerts (1989) develops a Methods of Simulated Moments estimation procedure for the parameters of option pricing models when Monte Carlo estimates substitute for intractable pricing formulae. By emphasizing conditional moments, Bossaerts avoids the complication of controlling both discretization and simulation errors.}

Duffie and Singleton's approach also differs from ours in that they focus on approximating the steady state moments rather than the conditional moments. One reason for focusing on the latter is that estimation using conditional moments is expected to lead to more efficient estimates than that using unconditional moments, since more precise information is used in forming the moment functions. However, the SME has the advantage of estimating a diffusion process with unobservable state variables. In our approach, we have to require that all of the state variables appearing in the diffusion equation be observable.

In order to develop large sample properties for the GMM estimators, we assume that the underlying diffusion process is stationary and ergodic. This allows us to draw upon laws of large numbers and central limit theorems, even though the sample observations are serially dependent. In the existing literature, many sets of conditions sufficient that ensure these large sample proper-

\footnote{Bossaerts (1989) develops a Methods of Simulated Moments estimation procedure for the parameters of option pricing models when Monte Carlo estimates substitute for intractable pricing formulae. By emphasizing conditional moments, Bossaerts avoids the complication of controlling both discretization and simulation errors.}
ties have been developed. For instance, the stationarity and ergodicity conditions can be replaced by a mixing condition. The mixing condition is always satisfied when the sample observations are generated from a finite Gaussian ARMA process. More generally, Galant and White (1988) employ the notion of near epoch dependence, which allows the serial observations to depend upon the entire history of another time series while still permitting the laws of large numbers and the central limit theorems to hold. Duffie and Singleton (1989) develop the notion of geometric ergodicity, a condition ensuring that the simulated process satisfies a law of large numbers with an asymptotic distribution that is invariant to the choice of initial conditions and the simulated process converges to the limiting distribution at a geometric rate. We adopt the stationarity and ergodicity assumptions which allow us to apply Hansen (1982)'s consistency and asymptotic normality theorems to establish the large sample properties of our GMM estimators.

As an illustration of our procedure, GMM estimation is applied to a mean reverting square root process originally used by Cox, Ingersoll and Ross (1985) for modeling the dynamics of the spot interest rate. In addition, we conduct Monte-Carlo simulations to examine the finite sample properties of the GMM estimators, focusing on the asymptotic normality of the estimator.

The paper is organized as follows. Section 2 sets out the GMM estimation procedure. Section 3 presents numerical approximation of the conditional moments for general diffusion processes. In Section 4, we develop the large sample properties for the GMM estimators introduced in Section 2. Section 5 reports the results of the GMM estimation applied to a mean reverting square root process, and simulations are conducted to examine the finite sample properties of the GMM estimators.

2 GMM Estimation

For simplicity, only univariate diffusion processes are considered here. The results can readily be extended to the general multi-dimensional case.

Let \( \{X(t) \in \mathbb{R} : t \in [0, \infty)\} \) be a stochastic process defined on a complete probability space \((\Omega, P, \mathcal{F})\). Suppose that \(X(t)\) satisfies the stochastic differential equation

\[
\frac{dX(t)}{dt} = b(X(t), \beta)dt + \sigma(X(t), \beta)dw(t),
\]

with \(X(0) = \tilde{Y}\). \(w\) is an one-dimensional standard Brownian motion defined on \((\Omega, P, \mathcal{F})\). \(\beta\) is an \(r\)-dimensional vector of unknown parameters to be estimated, and \(\beta\) lies in some compact parameter space \(\Theta \subset \mathbb{R}^r\). The true parameter, denoted by \(\beta_0\), is assumed to lie in the interior of \(\Theta\). The functions, \(b(x, \beta) : \mathbb{R} \times \mathbb{R}^r \to \mathbb{R}\) and \(\sigma(x, \beta) : \mathbb{R} \times \mathbb{R}^r \to \mathbb{R}\), are assumed to be continuous in both \(x\) and \(\beta\), and to satisfy the usual linear growth and uniform Lipschitz conditions. That is, there
exist constants $K, L > 0$ such that for all $x, y \in \mathbb{R}$ and $\beta \in \Theta$,

$$|b(x, \beta)| + |\sigma(x, \beta)| \leq K(1 + |x|),$$

(2)

$$|b(x, \beta) - b(y, \beta)| + |\sigma(x, \beta) - \sigma(y, \beta)| \leq L|x - y|.$$  

(3)

The above regularity conditions ensure that the solution to (1) exists and is unique. Note that $K$ and $L$ are assumed to be independent of $\beta$. This is reasonable given that $\Theta$ is compact. The process $X$ is time-homogenous, because the drift $b$ and the diffusion term $\sigma$ are independent of $t$. The time-homogeneity assumption is required for the stationarity of $X$, which we assume below. Finally, the starting point $\hat{Y}$ is a random variable.

A generalized method of moments approach to the estimation of the diffusion process can be conducted in the following manner. First, the process $X$ is sampled at dates $\{\tau_i\}$ for $t = 1, 2, \ldots, T$, yielding a series of equally spaced observations, denoted by $X(\tau_i)$ or $X_t$. Second, for each $t$, an $I$-dimensional vector valued measurable function of $X_t, X_{t-1}, \ldots, X_{t-k}$ and the unknown parameter $\beta$, denoted by $f_t(X_t, \ldots, X_{t-k}; \beta)$ or $f_t(X, \beta)$, (or $f_t^\beta$ for simplicity), is selected such that $f_t^\beta$ satisfies the following moment condition under the null hypothesis that $\beta = \beta_0$,

$$E[f_t(X, \beta; 0)] = 0.$$  

(4)

Usually, $I$ must be greater than or equal to $r$ in order for all of the unknown parameters to be identified. Lastly, for any parameter vector $\beta \in \Theta$, let

$$G_T(\beta) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t(X_t, \beta)$$

denote a vector of sample moments. If $f_t^\beta$ satisfies a strong or weak law of large numbers, then

$$\text{plim}_T G_T(\beta) = 0.$$  

With the usual identification assumptions, see Amemiya (1985), we have

$$\text{plim}_T G_T(\beta) = 0$$

if and only if $\beta = \beta_0$. The GMM estimator is obtained as the solution to the non-linear minimization program

$$\hat{\beta}_T \equiv \arg \min_{\beta \in \Theta} G_T(\beta)^T W_T G_T(\beta),$$

(5)

where $W_T$ is a symmetric, positive semi-definite $I \times I$ weighting matrix, measuring the length of the sample moments $G_T(\beta)$. Under the standard regularity conditions, $\hat{\beta}_T$ is consistent and asymptotically normal.

In general, the moment function, $f_t^\beta$, can be formed in two different ways. One way, as adopted by Duffie and Singleton (1989), is to use an unconditional expectations relation, which sets

$$f_t(X_t, \beta) = g_t(X_t, \beta) - E[g_t(X_t, \beta)],$$

(6)

\footnote{When $X$ is a stationary process, the distribution of $\hat{\beta}$ is the steady state distribution.}
where \( g_t = g(X_t, X_{t-1}, \ldots, X_{t-k}, \beta) \) is a known measurable function. The expectation operator \( E^\beta \) denotes the expectation taken as if the true parameter is \( \beta \). Alternatively, we can use a conditional expectations relation, which sets

\[
 f_t(X, \beta) = g_t(X, \beta) - E^\beta [g_t(X, \beta) | X_{t-l}],
\]

for some integer \( l > 0 \). In both cases, the moment conditions in (4) for \( f \) are satisfied.

One can see immediately that a complete knowledge of the functional form of \( E^\beta [g_t(X, \beta)] \) or \( E^\beta [g_t(X, \beta) | X_{t-l}] \) is essential for the minimization of (6) to be carried out. In general, it is very difficult to calculate the conditional moments of an arbitrary diffusion process. It is even more difficult to calculate the unconditional moments, since we may not know the distribution of \( \tilde{Y} \).

In this paper we focus on GMM estimation using the moment conditions derived from the conditional expectations relation as in (7). When the functional form of conditional moments is known, GMM estimation is straightforward. When this functional form is unknown, the numerical approximation procedure developed in the next section can be applied to approximate the conditional moments of the diffusion process. The basic idea of our approach is to approximate \( X \) by a sequence of finely spaced binomial processes (or multinomial processes if \( X \) is multi-dimensional) between \( \tau_{t-1} \) and \( \tau_t \), and replace the true conditional moments by the conditional moments calculated from the binomial (or multinomial) process. We shall demonstrate that the conditional moments obtained from the binomial (or multinomial) process converge uniformly to the true conditional moments of the limiting diffusion process.

Once the conditional moments are approximated, the GMM estimator can then be obtained as follows:

\[
 G_T^n(\beta) = \frac{1}{T} \sum_{t=1}^{T} f_t^n(X, \beta),
\]

\[
 \hat{\beta}_T^n = \arg \max_{\beta \in \Theta} G_T^n(\beta) W_T^n G_T^n(\beta),
\]

where \( f_t^n(X, \beta) \) is an approximation for \( f_t(X, \beta) \) and \( W_T^n \) is an approximation for the weighting matrix \( W_T \). If \( G_T^n(\beta) \) and \( W_T^n \) converge uniformly in probability to \( \text{plim}_T G_T(\beta) \) and \( \text{plim}_T W_T \) on \( \Theta \) as \( (T, n) \to \infty \), respectively, then it is reasonable to expect that \( \hat{\beta}_T^n \) has the same asymptotic distribution as \( \hat{\beta}_T \).

\[\text{Duffie and Singleton (1989) use a simulation technique to approximate the unconditional moments. Specifically, they set}
\]

\[
 E^\beta [g_t(X, \beta)] \approx \frac{1}{T} \sum_{t=1}^{T} g_t(\tilde{X}, \beta),
\]

where \( \tilde{X} \) is a simulated a sample path of the solution to the stochastic differential equation (1) and the length of the simulated sample path \( T \to \infty \).
3 Numerical Approximations of the Conditional Moments

We now describe how the conditional moments in (7) can be approximated by those conditional moments based on a sequence of binomial processes, where the sequence of binomial processes approximates the diffusion process in distribution. The idea of discrete time approximation originated in the contingent claims pricing literature, and in particular in Cox, Ross and Rubinstein (1979). Our objective here can be stated generally in terms of developing a numerical procedure to approximate expectations of functionals of the solution of the stochastic differential equation (1). We will provide sufficient conditions to ensure the discrete time moments converge uniformly to the actual continuous time moments.

For ease of notation, we assume that the time horizon is \([0, 1]\), and the conditional moment of interest has the following form:

\[
F(\beta) = E^\beta[g(X(1), \beta)|X(0)],
\]

where \(X(0)\) is fixed.\(^5\) The function \(g\) is a measurable mapping from \(\mathbb{R} \times \Theta\) to \(\mathbb{R}\) and satisfies a polynomial growth condition in the following sense: there exist constants \(K', \gamma > 0\) such that for all \(z \in \mathbb{R}\) and \(\beta \in \Theta\),

\[
|g(z, \beta)| \leq K'(1 + |z|^{\gamma}). \tag{9}
\]

We approximate the one dimensional diffusion process, \(X\), by a sequence of binomial processes in the following manner. The time interval \([0, 1]\) is divided into \(n\) equally spaced subintervals, each with a length of \(\frac{1}{n}\), and a binomial process \(X^n_k\) for \(k = 0, 1, \ldots, n\) is constructed as the solution to the stochastic difference equation

\[
\begin{align*}
X^n_{k+1} &= X^n_k + b(X^n_k, \beta) \frac{1}{n} + \sigma(X^n_k, \beta) \frac{\tilde{e}_k}{\sqrt{n}}, \\
X^n_0 &= X(0),
\end{align*} \tag{10}
\]

where \(P[\tilde{e}_k = 1] = P[\tilde{e}_k = -1] = \frac{1}{2}\). Equation (10) is a discretization of the stochastic differential equation (1).\(^6\)

Now, define function \(F^n\) as

\[
F^n(\beta) = E^n_0[g(X^n_0, \beta)], \tag{11}
\]

where \(E^n_0\) denotes the expectation under the probability measure for \(X^n_0\). Clearly, \(F^n\) is the conditional moment of \(X^n_0\), conditional on \(X^n_0\) being equal to \(X(0)\). Since the transition probabilities

\(^5\)Generalization to \(E^\beta[g(X(\tau_i), \beta)|X(\tau_{i-1})]\) is obvious.

\(^6\)Let \(X^n = X^n_{[n]}\). According to Theorem 7.4.1, Ethier and Kurtz (1986), \(\hat{X}^n\) converges to \(X\) weakly in \(D[0, 1]\), the space of functions from \([0, 1]\) to \(\mathbb{R}\) that are right-continuous with left limits.
for \( X_{k+1}^n \), conditional on \( X_k^n \), are constant and equal to \( \frac{1}{2} \), the expectation of \( \phi(X_n^n, \beta) \) conditional on \( X_0^n \) can easily be computed as

\[
F^n(\beta) = \frac{1}{2^n} \sum_{X_0^n} \phi(X_n^n, \beta).
\] (12)

Moreover, the derivatives of \( F^n \) with respect to \( \beta \) can be computed explicitly. Applying the chain rule, one can verify that

\[
\frac{\partial F^n(\beta)}{\partial \beta} = E^n_\beta \left[ \frac{\partial \phi(X_n^n, \beta)}{\partial \beta} D^n_\beta + \frac{\partial \phi(X_n^n, \beta)}{\partial \beta} \right],
\]

where \( D^n_\beta = \frac{\partial X^n}{\partial \beta} \) is an \( r \)-dimensional vector of the derivative process of \( X_k^n \) with respect to \( \beta \), and \( D^n_\beta \) satisfies the stochastic difference equation

\[
D^n_{k+1} = D^n_k + (b_\beta(X_k^n, \beta) + b_x(X_k^n, \beta) D^n_k) \frac{1}{n} + (\sigma_\beta(X_k^n, \beta) + \sigma_x(X_k^n, \beta) D^n_k) \frac{\bar{\epsilon}_k}{\sqrt{n}},
\]

\[
D^n_0 = 0,
\]

where \( b_\beta, b_x, \sigma_\beta \) and \( \sigma_x \) denote the partial derivatives of \( b(x, \beta) \) and \( \sigma(x, \beta) \) with respect to \( x \) and \( \beta \), respectively.\(^7\) The following proposition establishes the uniform convergence of \( F^n \) and \( \frac{\partial F^n}{\partial \beta} \) to \( F \) and \( \frac{\partial F}{\partial \beta} \), respectively.

**Proposition 1** Suppose:

a) \( b \) and \( \sigma \) are continuously differentiable with respect to \( x \) up to the seventh orders; \( b_\beta \) and \( \sigma_\beta \) exist and are continuously differentiable with respect to \( x \) up to the sixth order, and these derivatives satisfy polynomial growth conditions (with possibly different \( K' \) and \( \gamma \), independent of \( \beta \));

b) \( g \) and \( g_\beta \) are continuously differentiable up to the seventh and the sixth orders, respectively, and all these derivatives satisfy a polynomial growth condition as in a);

c) \( b_x(x, \beta) y, \sigma_x(x, \beta) y, b_\beta(x, \beta) \) and \( \sigma_\beta(x, \beta) \) satisfy uniform Lipschitz conditions in \( x \) and \( y \) respectively (with possibly different \( K \), independent of \( \beta \)).

Then, there exist constants \( C \) and \( \gamma' > 0 \), independent of \( n \) and \( \beta \), such that, for all \( \beta \in \Theta \),

1) \( |F^n(\beta) - F^n(\beta)| \leq C \frac{\gamma'}{n} (1 + |X_0|^{\gamma'}) \),

2) \( |\frac{\partial F^n(\beta)}{\partial \beta} - \frac{\partial F^n(\beta)}{\partial \beta}| \leq C \frac{\gamma'}{n} (1 + |X_0|^{\gamma'}) \).

**Proof.** See Appendix.

\(^7\)More generally, following Gihman and Skorohod (1975), we have \( \frac{\partial F^n(\beta)}{\partial \beta} = E^n_\beta \left[ \frac{\partial g(X(t), \beta)}{\partial \beta} D(t) + \frac{\partial g(X(t), \beta)}{\partial \beta} \right], \) where \( D \) is the derivative process of \( X \) w.r.t. \( \beta \), which satisfies the equation

\[
dD(t) = (b_\beta(X(t), \beta) + b_x(X(t), \beta) D(t)) dt + (\sigma_\beta(X(t), \beta) + \sigma_x(X(t), \beta) D(t) D(t)) dw(t).
\]
Remark 1 i) Condition c) can be relaxed to requiring only that $b(x, \beta)$ and $\sigma(x, \beta)$ satisfy a uniform Lipschitz condition in $x$. However, the stronger assumptions of c) greatly simplify the proof of 2).

ii) Result 1) will be used to establish consistency, while result 2) will be used to establish asymptotic normality.

Proposition 1 shows that the discretization scheme defined by stochastic difference equation (10) produces approximations for conditional moments of the order $\frac{1}{n}$. Milshtein (1974) develops a discretization scheme that produces approximations for conditional moments of the order $\frac{1}{n^2}$. However, when $X$ is multi-dimensional, the Milshtein scheme requires additional assumptions on the diffusion term $\sigma$ which may not be satisfied. We emphasize that in general the number of discrete states in binomial approximation grows geometrically with respect to the number of time steps since a state with a subsequent upward and downward move does not match a state with a subsequent downward and upward move. When $X$ is one-dimensional, it is possible to construct a binomial approximation where the number of discrete states grows linearly (see Nelson and Ramaswamy (1989)).

The discrete time approximation of $X$ when it is multi-dimensional can be dealt with similarly. This involves approximating the $N$-dimensional Brownian motion by a sequence of multinomial processes. For example, a multinomial process formed by putting $N$ independent binomial processes together comprises an approximation for the $N$-dimensional Brownian motion.\(^8\)

Finally, to obtain an approximation for the objective function, the above discretization procedure is applied repeatedly to all of the consecutive sample dates to yield conditional moments approximations for $E\beta[g_t(X, \beta)|X_{t-1}]$, $t = 1, 2, \cdots, T$. Once $F$ is approximated by $F^n$, we can then replace $f_t$ by

$$f^n_t = g_t(X, \beta) - F^n(X_t, \beta),$$

and apply the GMM estimation procedure to get the GMM estimator $\hat{\beta}_T^n$ defined in (8).

4 Large Sample Properties

In this section we establish the large sample properties of the GMM estimators $\hat{\beta}_T$ and $\hat{\beta}_T^n$ defined in Section 2. The basic treatment follows closely that of Hansen (1982).

\(^8\)For example, if $N = 2$, we can use $\xi^1_t, \xi^2_t$ to approximate the local increments of two Brownian motions, where

$$P[\xi^1_t = 1, \xi^2_t = 1] = P[\xi^1_t = 1, \xi^2_t = -1] = P[\xi^1_t = -1, \xi^2_t = 1] = P[\xi^1_t = -1, \xi^2_t = -1] = \frac{1}{4}.$$
4.1 GMM Estimation when \( f_t^\beta \) is Known.

To establish the large sample properties, we need to make some basic assumptions on the maximum degree of dependence allowed between the discretely sampled observations \( \{X_t\}_{t=1}^\infty \) which will still permit some form of the law of large numbers and the central limit theorem to obtain.\(^9\) The standard results require that \( \{X_t\} \) be stationary and ergodic, as in Billingsley (1969) and Hansen (1982). Mcleish (1975) shows that one could replace the stationarity and ergodicity assumptions by a mixing condition (for an elementary discussion of the mixing condition, see White (1988)). For stationary sequences, the mixing condition is stronger than ergodicity, since it implies ergodicity. More recently, Galant and White (1988) employ the notion of near epoch dependency, which generalizes the notion of mixing to a time series, the observations of which may depend upon the entire history of another time series. In practice, both mixing and ergodicity conditions are difficult to verify. Duffie and Singleton (1989) take a different approach and introduce the notion of geometric ergodicity, for which easily verifiable sufficient conditions exist, although they may be too strong to apply in practice.\(^10\) Intuitively, geometric ergodicity implies that the stochastic process is asymptotically stationary and ergodic. For the purpose of this paper, any of the above mentioned assumptions suffices. We adopt the stationarity and ergodicity assumptions, which would then allow us to apply Hansen (1982)'s theorems regarding the consistency and asymptotic normality of the GMM estimators.

We first summarize here several important regularity conditions for consistency.

**Assumption 1 (Ergodicity)** \( \{X_t, t = 1, 2, \ldots, \} \) is stationary and ergodic.

When the diffusion process \( \{X(t), t \in [0, \infty)\} \) is stationary and ergodic, the sequence of discretely sampled observations \( \{X_t, t = 1, 2, \ldots, \} \) is also stationary and ergodic. Hence, \( \{f_t^\beta, t = 1, 2, \ldots, \} \) is stationary and ergodic for all \( \beta \in \Theta \).

**Assumption 2 (Weighting Matrix)**

\[
\mathbf{W}_T \xrightarrow{p} \mathbf{W}_0,
\]

\(^9\)In cases where the diffusion process has a long-run growth trend, \( X \) may be a transformation of the original process such that \( X \) is trend-free.

\(^10\)Formally, let \( P_t^\gamma \) denote the transition probability of \( X(t) \) given the initial point \( X(0) = x \) for a time-homogeneous Markov process \( \{X(t)\} \). The process \( \{X(t)\} \) is called geometrically ergodic if there exists some \( \rho \in (0, 1) \), and a probability measure \( \pi \) on the state space of the process such that, for every initial point \( x \),

\[
\rho^{-t} ||P_t^\gamma - \pi||_\pi \to 0, \text{as } t \to \infty,
\]

where \( || \cdot ||_\pi \) denotes the total variation norm. Duffie and Singleton (1989) demonstrate that in order for a diffusion process to be geometrically ergodic, the drift of the diffusion must be negative. They also argue that geometric ergodicity implies \( \alpha \)-mixing.
where $W_0$ is an $I \times I$, non-stochastic, and positive semi-definite matrix with $\text{rank}(W_0) \geq r$.

The weighting matrix $W_T$ defines a measure of the length of the sample moment $G_T(\beta)$. As suggested in Hansen (1982), one may choose $W_0 = S_0^{-1}$ to get an efficient GMM estimator, where

$$S_0 = \sum_{j=-\infty}^{\infty} E(f_{t-j}^\beta f_{t-j}^{\beta^\top}).$$

In our setup, since the moment conditions are formed by conditional expectations, $\{f_{t-j}^\beta\}$ is a martingale difference sequence, which implies that $E(f_{t-j}^\beta f_{t-j}^{\beta^\top}) = 0$ for $j \neq 0$. Hence, $S_0$ is simply equal to $E[f_{t}^\beta f_{t}^{\beta^\top}]$.

**Assumption 3 (Regularity Condition I)**

1. $f_t(X, \beta)$ is measurable in $X$ for each $\beta \in \Theta$ and continuous in $\beta$ for fixed $X$;
2. $Ef_t(X, \beta)$ exists and is finite for all $\beta \in \Theta$;
3. $f_t(X, \beta)$ is first moment continuous at $\beta$ for all $\beta \in \Theta$;
4. $H(\beta) \equiv Ef_t^\top(X, \beta)W_0Ef_t(X, \beta)$ has a unique minimum at $\beta_0$.

Formally, a random function $f(\omega, \beta)$ is first moment continuous at $\beta$ if

$$\lim_{\delta \to 0} Ef[\epsilon(\omega, \beta, \delta)] = 0,$$

where $\epsilon(\omega, \beta, \delta) = \sup\{|f(\omega, \beta) - f(\omega, \alpha)| : \alpha \in \Theta, |\beta - \alpha| < \delta\}$. A sufficient condition for $f$ to be first moment continuous is that there exists a $\delta > 0$ such that $Ef[\epsilon(\omega, \beta, \delta)] < \infty$ (see Lemma 2.1, Hansen (1982)). As a consequence, $f$ is first moment continuous if

$$E \sup_{\beta \in \Theta} f_{t}^\beta < \infty. \quad (13)$$

The first moment continuity condition ensures that when we apply the weak law of large numbers to $\{f_{t}^\beta\}$, the sample moments converge in probability to their population moments uniformly in $\beta$.\(^{11}\)

Assumptions 1, 2 and 3 are standard assumptions for consistency. The following theorem summarizes the consistency result for the GMM estimator $\hat{\beta}_T$ in the case where the functional form of conditional moments is known. The proof of this theorem follows directly from Theorem 2.1, Hansen (1982).

\(^{11}\)Stronger regularity conditions such as Lipschitz continuity (or smoothness) are required for the uniform law of large numbers (ULLN's) to hold, if $X$ is mixing or geometrically ergodic, but not stationary (see Andrew (1987), Duffie and Singleton (1989) and Potscher and Prucha (1989)).
Theorem 1 (Consistency for $\hat{\beta}_T$) If Assumptions 1 through 3 are satisfied, then $\hat{\beta}_T$ exists and converges in probability to $\beta_0$.

We need two more assumptions for deriving the asymptotic distribution of $\hat{\beta}_T$.

Assumption 4 (Regularity Condition II)

(i) The estimators $\hat{\beta}_T$ lie in the interior of $\Theta$;
(ii) $f_t^\beta$ is continuously differentiable with respect to $\beta \in \Theta$ for all $t$;
(iii) $\partial f_t^\beta / \partial \beta$ is first moment continuous at $\beta_0$;
(iv) $d_0 = E[\partial f_t^{\beta_0} / \partial \beta]$ exists, and has full rank.

Assumption 5 $S_0 = E[f_t^{\beta_0} f_t^{\beta_0\top}]$ exists and is invertible. Moreover, $\text{plim}_T W_T = S_0^{-1}$.

The following proposition establishes the asymptotic distribution of $\hat{\beta}_T$, the proof of which also follows directly from Theorem 3.1, Hansen (1982).

Theorem 2 (Asymptotic Normality for $\hat{\beta}_T$) Under Assumptions 1 through 5, $\sqrt{T}(\hat{\beta}_T - \beta_0) \rightarrow N(0, D)$ as $T \rightarrow \infty$, where $D = (d_0^\top S_0^{-1} d_0)^{-1}$.

4.2 GMM Estimation when $f_t$ is Unknown

We now investigate the large sample properties of the GMM estimator $\hat{\beta}_T^n$ in the case where the functional form of the conditional moments is unknown and the actual conditional moments are replaced by their discrete time approximations. We shall show that under certain regularity conditions, $\hat{\beta}_T^n$ has the same large sample distribution as $\hat{\beta}_T$ when $T$ and $n$ go to infinity.

We need to make a similar set of regularity assumptions.

Assumption 2' (Weighting Matrix)

$$\text{p}_{\lim T \rightarrow \infty} W_T^n = W^n,$$

where $W^n$ is an $I \times I$, non-stochastic, and positive semi-definite matrix with rank($W^n$) $\geq r$. In addition,

$$\| W^n - W_0 \| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

where $W_0$ is defined as in Assumption 2, and $\| A \| \equiv (\text{trace}(A^\top A))^{1/2}$ denotes the norm of matrix $A$.

Assumption 3' (Regularity Condition I')
i) $f^n_\beta(X, \beta)$ is measurable in $X$ for each $\beta \in \Theta$ and continuous in $\beta$ for fixed $X$;

ii) $Ef^n_\beta(X, \beta)$ exists and is finite for all $\beta \in \Theta$;

iii) $I^n_\beta(X, \beta)$ is first moment continuous at $\beta$ for all $\beta \in \Theta$;

iv) $H^n(\beta) \equiv Ef^n_\beta(X, \beta)W^nEf^n_\beta(X, \beta)$ has a unique minimum at $\beta^n$, where $\beta^n$ lies in the interior of $\Theta$.

To understand Assumption 3', we introduce the following lemma, which is also important for later theorems.

**Proposition 2** Let

$$R^n(\beta) = E[f^n_\beta(X, \beta)], \quad R(\beta) = E[f(I(X, \beta))]$$

$$J^n(\beta) = E[\frac{\partial f^n_\beta(X, \beta)}{\partial \beta}], \quad J(\beta) = E[\frac{\partial f(I(X, \beta))}{\partial \beta}]$$

$$H^n(\beta) = E[f^n_\beta(X, \beta)]' W^n E[f^n_\beta(X, \beta)], \quad H(\beta) = E[f(I(X, \beta))]' W_0 E[f(I(X, \beta))]$$

Suppose that the assumptions of Proposition 1 are satisfied and that $E|Y|^{\lambda} < \infty$ for any $\lambda > 0$. Then $R^n$, $J^n$, and $H^n$ converge uniformly on $\Theta$ to $R$, $J$ and $H$, respectively.

**PROOF.** See Appendix.

Given this proposition, conditions ii) and iii) of Assumption 3' can easily be derived from conditions ii) and iii) of Assumption 3. Since we have assumed that $H(\beta)$ has a unique minimum at $\beta_0$, condition iv) of Assumption 3' is reasonable to impose, given that $H(\beta)$ converges uniformly to $H^n(\beta)$. We now present the following consistency result for $\hat{\beta}^n_\pi$.

**Theorem 3** (Consistency for $\hat{\beta}^n_\pi$

Under Assumptions 1, 2, 3 and the assumptions of Proposition 2, $\hat{\beta}^n_\pi$ defined in (3) is consistent as $(n, T)$ goes to infinity in the following manner. For any $\epsilon, \delta > 0$, there exist $n_0 > 0$ and $T(n) > 0$ for $n > n_0$, such that

$$P(\|\hat{\beta}^n_\pi - \beta_0\| > \delta) < \epsilon,$$

for all $n > n_0$ and $T > T(n)$.

**PROOF.** Under the given assumptions, $G^n_\pi(\beta)' W^n_0 G^n_\pi(\beta) \rightarrow H^n(\beta)$ in probability uniformly on $\Theta$ for fixed $n$ as $T$ goes to infinity. This implies that $\hat{\beta}^n_\pi \rightarrow \beta^n$ in probability as $T \rightarrow \infty$, see Amemiya (1985).

According to Proposition 2, $H^n$ converges to $H$ uniformly, which implies that $\beta^n \rightarrow \beta_0$ as $n \rightarrow \infty$. Thus, for any $\delta > 0$, there exists $n_0 > 0$ such that $|\beta^n - \beta_0| < \delta/2$ for all $n > n_0$. 

14
Now, for any $\epsilon > 0$, since $\text{plim}_{T} \hat{\beta}_T^n = \beta^n$, we can find a $T(n) > 0$ such that as $T > T(n)$, $P[|\hat{\beta}_T^n - \beta^n| > \delta/2] < \epsilon$. This implies that

$$P[|\hat{\beta}_T^n - \beta_0| > \delta] \leq P[|\hat{\beta}_T^n - \beta^n| > \delta/2] < \epsilon$$

for all $n > n_0$ and $T > T(n)$.

It is clear from the preceding proof that, for fix $n$, $\hat{\beta}_T^n$ might be inconsistent even when $T$ goes to infinity. The bias is $\beta_0 - \beta^n$, which can be very small if the approximation is good enough. We have to let both $n$ and $T$ go to infinity in order for $\hat{\beta}_T^n$ to be consistent. We make some additional regularity assumptions, which will yield an asymptotic distribution for $\hat{\beta}_T^n$.

**Assumption 4’ (Regularity Condition II’)**

(i) The estimators $\hat{\beta}_T^n$ lie in the interior of $\Theta$;

(ii) $f_t^n(X, \beta)$ is continuously differentiable with respect to $\beta \in \Theta$ for all $t$;

(iii) $\partial f_t^n(X, \beta)/\partial \beta$ is first moment continuous at $\beta^n$;

(iv) $d^m = E[\partial f_t^n(X, \beta^n)/\partial \beta]$ exists, and has full rank.

**Assumption 5’** $S^n = E[f_t^n(X, \beta^n)f_t^n(X, \beta^n)^\top]$ exists and is invertible. Moreover, $\text{plim}_{T} W_T^n = S^{n-1}$.

The following theorem derives the asymptotic distribution for the GMM estimator $\hat{\beta}_T^n$.

**Theorem 4** (Asymptotic Normality for $\hat{\beta}_T^n$) Under Assumptions 1, 2 through 5’ and the assumptions of Proposition 2, $\sqrt{T}(\hat{\beta}_T^n - \beta_0) \rightarrow N(0, D)$ as $(n, \infty) \rightarrow \infty$, where $D = (d^\top S^{-1} d)^{-1}$.

**PROOF.** Following Theorem 3.1, Hansen (1982), we know that for fix $n$,

$$\sqrt{T}(\hat{\beta}_T^n - \beta^n) \rightarrow N(0, (d^n \top S^n d^n)^{-1}), \quad \text{as } T \rightarrow \infty,$$

where $d^n = E^{\partial f_t^n(X, \beta^n)/\partial \beta}$.

According to Proposition 2, $E^{\partial f_t^n(X, \beta)/\partial \beta}$ converges to $E^{\partial f_t^n(X, \beta)/\partial \beta}$ uniformly on $\Theta$. Letting $W^n = W_0 = I$ in Proposition 2, we know that $E[f_t^n(X, \beta)f_t^n(X, \beta)^\top]$ converges to $E f_t(X, \beta)f_t(X, \beta)^\top$ uniformly on $\Theta$. We conclude that $d^n \rightarrow d_0$ and $S^n \rightarrow S_0$ as $n \rightarrow \infty$. Hence,

$$\sqrt{T}(\hat{\beta}_T^n - \beta_0) \rightarrow N(0, D),$$

in distribution as $(n, T) \rightarrow \infty$ as in Theorem 3.

This concludes our demonstration of the large sample properties of the GMM estimators.
5 Example: Square Root Process

In this section, we apply the proposed GMM estimation to a mean reverting square root process originally used by Cox, Ingersoll and Ross [1985] to model the dynamics of the spot interest rate. We present the results of several sampling experiments which explore the finite sample properties of the GMM estimators \( \hat{\beta}_T \) and \( \hat{\beta}_T^p \) developed in the previous sections. In particular, these experiments examine whether the finite distributions of \( \hat{\beta}_T \) and \( \hat{\beta}_T^p \) are close to their theoretical asymptotic normal distributions as \( T \) and \( n \) become large.

Formally, the mean reverting square root process \( \{X(t), t \geq 0\} \) is determined by the stochastic differential equation

\[
dX(t) = \kappa(\theta - X(t)) + \sigma \sqrt{X(t)} \, dw(t),
\]

where \( \theta \) is the long-run mean and \( \kappa \) is the intensity of the mean reversion. We assume that \( \kappa, \theta > 0 \) and \( 2\kappa\theta \geq \sigma^2 \), which together imply that zero is inaccessible. The square root process has a stationary distribution with the density function

\[
f(x) = \frac{\omega^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\omega x}, \quad x \geq 0,
\]

where \( \omega = \frac{2\kappa}{\sigma^2} \), \( \nu = \frac{2\kappa\theta}{\sigma^2} \), and \( \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \). It is known that the mean reverting square root process \( \{X(t), t \geq 0\} \) is stationary and ergodic if zero is inaccessible.\(^1\) The transition density function of \( X(s) \) given \( X(t) \), when \( s > t \), is given by

\[
f(X(s), s \mid X(t), t) = ce^{-u-v} \left( \frac{v}{u} \right)^{\frac{q}{2}} I_q(2uv^{\frac{1}{2}}),
\]

where

\[
c \equiv \frac{2\kappa}{\sigma^2(1 - e^{-\kappa(s-t)})},
\]

\[
u \equiv cX(t)e^{-\kappa(s-t)},
\]

\[v \equiv cX(s),
\]

\[
q \equiv \frac{2\kappa\theta}{\sigma^2} - 1,
\]

and \( I_q(\cdot) \) is the modified Bessel function of the first kind of order \( q \).

\([^1\)One way to prove ergodicity is to show that there exists a constant \( \lambda > 0 \) such that

\[\| P_t^s - \pi \|_s \leq \lambda e^{-\lambda t},\]

where \( \pi \) is the steady state probability measure on the state space. This implies that \( X \) is geometrically ergodic, and therefore is \( \alpha \)-mixing, according to Duffie and Singleton (1989). Mixing implies ergodicity.\]
We conduct sampling experiments in the following manner. First, for a fixed set of parameter values for $\kappa, \sigma$ and $\theta$, a random sample of weekly observations of $X$ is generated. Second, a GMM estimation is performed, which yields the point estimates and the estimated variance-covariance matrix for $\kappa, \sigma$ and $\theta$. This sampling-estimation procedure is then repeated independently 500 times, which produces 500 replications of the sampling experiment. Since the mean reverting intensity parameter, $\kappa$, has a strong effect on the pattern of the sample path of $X$, three sets of parameters ($\kappa = 0.8, 4, 40$) are used in these experiments, while $\theta$ and $\sigma$ are held constant at 0.1. The effect of mean reversion intensity on the statistical properties of point estimates will be explored below.

In order to generate a random sample of size $T$ for $X$, the stochastic differential equation (1) is discretized and simulated in the following way,

$$
\tilde{X}_{k+1}^h = \tilde{X}_k^h + \kappa(\theta - \tilde{X}_k^h)\Delta t + \sigma \sqrt{\tilde{X}_k^h} \sqrt{\Delta t} \tilde{z}_k,
$$

where $h$ is set to be $1/(52 \times 200)$, $\tilde{z}_k$ is a sequence of i.i.d. random variables and these random variables have normal $N(0, 1)$ distributions. The initial value, $X(0)$, is drawn from the stationary distribution as in (15). The sample observations, $(X_1, X_2, \ldots, X_T)$, are then obtained by setting $X_t = \tilde{X}_{200(t-1)}^h, \quad t = 1, 2, \ldots, T$.

The discretization procedure in (16) may involve errors causing the simulated observations $X_t$ to differ from a random sample drawn from the square root process. However, it is reasonable to believe that such effect is insignificant, since it is known that $\tilde{X}_k^h$ converges in distribution to the square root process as $h \to 0$.

Experiments are performed for sample sizes of $T = 260, 520, 780$ and 1040, which correspond in real time to series of weekly observations of lengths 5, 10, 15 and 20 years, respectively. Since there are three unknown parameters, we need at least three moment conditions. We used the following three conditions

$$
\begin{align*}
\ell_1^1(\beta) &= X_{t+1} - E^\beta[X_{t+1}|X_t], \\
\ell_2^1(\beta) &= X_{t+1}^2 - E^\beta[X_{t+1}^2|X_t], \\
\ell_2^2(\beta) &= (X_{t+1} - E^\beta[X_{t+1}|X_t])^2 = \frac{\text{Var}^\beta(X_{t+1}|X_t)}{\text{Var}^\beta(X_{t+1}|X_t)} - 1,
\end{align*}
$$

where $\beta = (\kappa, \sigma, \theta)^T$,

$$
\begin{align*}
E^\beta[X_{t+1}|X_t] &= X_t e^{-\kappa \Delta} + \theta (1 - e^{-\kappa \Delta}), \\
E^\beta[X_{t+1}^2|X_t] &= X_t \frac{\sigma^2}{\kappa} (e^{-\kappa \Delta} - e^{-2\kappa \Delta}) + \frac{\sigma^2}{2\kappa} (1 - e^{-\kappa \Delta})^2 + (X_t e^{-\kappa \Delta} + \theta (1 - e^{-\kappa \Delta}))^2,
\end{align*}
$$

17
\[ \text{Var}^\beta(X_{t+1}|X_t) = X_t \sigma^2 (e^{-\kappa \Delta} - e^{-2\kappa \Delta}) + \theta \sigma^2 (1 - e^{-\kappa \Delta})^2, \]

and \( \Delta = 1/52 \), which gives a time interval of one week. It is not difficult to verify that Assumptions 1 through 5 are satisfied, as long as \( \kappa \), \( \sigma \) and \( \theta \) are assumed to be positive and lie in a compact set that does not contain 0. Now, define

\[ G_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} f_t(X_t, \beta). \]

Our estimation procedure is carried out by performing the following minimization over \( \beta \):

\[ \min_{\beta} G_T(\beta)^\top G_T(\beta). \]

Since the parameters are exactly identified, the optimal solution is obtained when \( \hat{\beta}_T \) is such that \( G(\hat{\beta}_T) = 0 \). In order to calculate the variance-covariance matrix, we define \( \hat{W}_T \equiv \Sigma^{-1}(\hat{\beta}_T) \), where

\[ \Sigma(\hat{\beta}_T) = \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\beta}_T)f_t(\hat{\beta}_T)^\top. \]

The asymptotic variance-covariance matrix for \( \hat{\beta}_T \) is then calculated as

\[ \text{Var}(\hat{\beta}_T) = \frac{1}{T} \left[ D(\hat{\beta}_T) \hat{W}_T D(\hat{\beta}_T) \right]^{-1}, \tag{17} \]

where

\[ D(\hat{\beta}_T) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f_t^\beta}{\partial \beta} \bigg|_{\beta = \hat{\beta}_T}. \]

Tables 1, 2, and 3 summarize the basic results of the simulations for \( \kappa = 0.8 \) with \( T = 520,780,1040 \) and \( \kappa = 4,40 \) with \( T = 260,620,780 \). Each table corresponds to a separate independent experiment involving 500 independent replications. The second column displays the values of the true parameters. The third column displays the means and standard deviations of the point estimates of \( \hat{\beta}_T \). The fourth column reports the theoretical values for the asymptotic variance and standard deviation of the estimators \( \hat{\beta}_T \), while the fifth column reports the means and standard deviations of the estimated asymptotic variance calculated according to (17). The next three columns display estimated 1, 5 and 10 percent tail probabilities, respectively.\(^{14}\)

Asymptotic estimation is performed in two stages. In the first stage, a consistent estimator is found by using an arbitrary weighting matrix. This consistent estimate is then used to form an optimal weighting matrix to yield an efficient consistent estimator. However, when the number of moment conditions is exactly equal to the number of unknown parameters, any weight matrix will yield the same asymptotic variance-covariance matrix for \( \hat{\beta}_T \).

\(^{14}\)For given size \( \alpha \), let \( A_\alpha \) be such that

\[ P\left[ \frac{\kappa_T - \kappa_0}{\sigma(\kappa_T)} > A_\alpha \right] = \alpha, \]

where \( \kappa_T \sim N(\kappa_0, \sigma(\kappa_T)^2) \), then the tail probability for \( \kappa_T \) is defined to be \( \frac{m}{500} \), where \( m \) is the total number of times among 500 replications that \( \kappa_T \) lies outside of \( (\kappa_0 - \sigma(\kappa_T)A_\alpha, \kappa_0 + \sigma(\kappa_T)A_\alpha) \). The tail probabilities for \( \hat{\sigma}_T \) and \( \hat{\theta}_T \) are defined similarly.

18
$t$-statistics for the hypotheses that the true tail probabilities are 1, 5 and 10 percent respectively are reported in parentheses. The last column reports the Pearson $\chi^2$-statistics and their probability values under the null hypothesis that $\hat{\beta}_T$ has an (asymptotic) normal distribution.\footnote{We divide the real line into 20 intervals such that each interval has a probability of 0.05 under the null hypothesis. The Pearson $\chi^2$-statistics is constructed as

$$\chi^2 = \sum_{i=1}^{20} \frac{(f_i - 0.05 \times 500)^2}{0.05 \times 500},$$

where $f_i$ is the frequency of the $i$-th interval. The probability value is calculated as $P[\chi^2(19) \geq \chi^2]$.}

The results reported in Tables 1, 2 and 3 suggest that the point estimates for $\sigma$ and $\theta$ are approximately consistent in all nine experiments. As we would have expected, the standard deviations decrease as the sample size $T$ increases. While the standard deviations for $\hat{\sigma}_T$ in column three are largely unchanged for three different levels of $\kappa$, the standard deviations for $\hat{\theta}_T$ decrease as $\kappa$ increases. Moreover, the standard deviations of point estimates in column three are close to their theoretical values for the asymptotic standard deviation in column four, while the means of the estimated asymptotic variances in column five are also close to their theoretical values for the asymptotic variance in column four. In addition, the $t$-statistics for the tail probabilities as well as the $\chi^2$ probability values are largely insignificant. These results suggest that the finite sample distributions of $\hat{\sigma}_T$ and $\hat{\theta}_T$ match the asymptotic normal distributions established in Section 4 reasonably well.

The finite sample properties of the point estimates for $\kappa$ are much worse than those of $\sigma$ and $\theta$. When the true $\kappa$ is 0.8, the point estimates for $T = 520, 780$ and 1040 are strongly biased, and the tail probabilities as well as the $\chi^2$ statistics are significantly different from their corresponding true values, although the point estimate for $T = 1040$ improves quite a bit. It is clear from Table 1 that when the true $\kappa$ is equal to 0.8, a longer time series is needed in order to obtain good estimates for $\kappa$. As the true $\kappa$ increases to 4 and 40, the finite sample properties of the point estimates improve significantly, and point estimates for $\kappa$ are approximately consistent for $T \geq 520$. The $t$-statistics for the tail probabilities and $\chi^2$ probability values for $T \geq 520$ are largely insignificant, and the means and standard deviations of the point estimates are close to their corresponding theoretical values, which suggest that the finite distributions of $\hat{\sigma}_T$ resemble their theoretical asymptotic normal distributions quite well.

Overall, the results strongly suggest that it is much easier to estimate $\kappa$ when the true $\kappa$ is large than when it is small. This is not surprising. When the true $\kappa$ is small, the sequence looks like a random walk or a unit root process. Thus it is difficult to identify $\kappa$. In fact, with a small sample, the estimator for $\kappa$ exhibits an upward bias (see Ball (1989)).
We now consider the GMM estimator $\hat{\beta}_T^n$, where we replace $f^1$, $f^2$ and $f^3$ by the conditional moments calculated from (12). Although the regularity conditions required for Proposition 1 are not satisfied for the square root process, we can prove directly that the mean and variance obtained from the binomial process will produce approximations for the actual mean and variance of the order $\frac{1}{n}$. As illustrations, we compare the values and derivatives of the actual conditional moments and the actual objective function to their corresponding approximations. The results are reported in Tables 4, 5, 6 and 7, where we have fixed the parameters at $\kappa = 0.8, 4, 40$ and $\sigma = \theta = 0.1$. Clearly, from these tables, the approximations for the first three moments and the objective function are quite good, with $\kappa = 0.8$ being the best.

Finally, we apply GMM estimation for $\kappa = 4$, $T = 520$, and $n = 12$ to 10 independent sample observations. The results of comparisons between $\hat{\beta}_T$ and $\hat{\beta}_T^n$ are reported in Table 8, where each row corresponds to one pair of estimates obtained by using the same sequence of time series observations of size 520. The point estimates of $\hat{\beta}_T^n$ and $\hat{\beta}_T$ and the estimated standard deviations are virtually identical. We conclude that $\hat{\beta}_T^n$ has the same asymptotic distribution as $\hat{\beta}_T$.

6 Concluding Comments

We have developed an approach for estimating the parameters of general diffusion processes using GMM. Our procedure involves numerical approximation of the conditional moments when the true functional form of the conditional moments are not available. This procedure is easy to apply in practice. We observe that when a diffusion process is approximated by a binomial (multinomial) process, the probability distribution of the diffusion process can be approximated by that of the binomial (multinomial) process. This suggests that a maximum likelihood estimation may be applicable. ML estimation is important since it will produce an efficient estimator if the parametric specification of the diffusion process is correct. The main difficulty in applying ML estimation is the construction of a continuous and differentiable density function that approximates the true density function.

\footnote{Since the conditional mean and the conditional second moment are linear and quadratic functions respectively, we need only to show that the inequality for $S_T^n$ in Lemma 1 is satisfied. This can be checked easily.}
References


Appendix. Proofs

We need a lemma to proceed with the proof for Proposition 1.

Lemma 1 Suppose that assumptions a), b) and c) of Proposition 1 are satisfied. Then, for any integer \( m, k \geq 0 \) and \( k \leq n \), there exists a constant \( A > 0 \), dependent only upon \( m, K \) and \( L \), such that

\[
E|X(t)|^{2m} \leq A(1 + E|X(0)|^{2m}),
\]

and

\[
E|X^n_t|^{2m} \leq A(1 + E_n|X^n_0|^{2m}).
\]

Proof. The first inequality follows Theorem 2.3 of Friedman (1975, pp. 107). We prove the second inequality below. The proof follows closely the proof of Theorem 2.3 of Friedman.

We can find some constant \( K' > 0 \), depending only upon \( K \) and \( L \), such that for any \( x \in \mathbb{R} \),

\[
|b(x)| \leq K'(1 + |x|), \quad |b(x)|^2 \leq K'(1 + |x|^2), \quad |\sigma(x)| \leq K'(1 + |x|),
\]

\[
|\sigma(x)|^2 \leq K'(1 + |x|^2), \quad |xb(x)| \leq K'(1 + |x|^2), \quad |b(x)\sigma(x)| \leq K'(1 + |x|^2),
\]

Applying a Taylor expansion to the function \( x^{2m} \), we obtain

\[
|X^n_{k+1}|^{2m} = |X^n_k|^{2m} + 2m[X^n_k]^{2m-1}(X^n_{k+1} - X^n_k) + m(2m - 1)|X^n_k|^{2m-2}(X^n_{k+1} - X^n_k)^2
\]

\[
= |X^n_k|^{2m} + 2m[X^n_k]^{2m-1}\left(\frac{b}{n} + \frac{\sigma}{\sqrt{n}}\epsilon_k\right) + m(2m - 1)|X^n_k|^{2m-2}\left(\frac{\hat{b}}{n^2} + \frac{2b\sigma}{n\sqrt{n}}\epsilon_k + \frac{\sigma^2}{n}\epsilon_k^2\right),
\]

where \( \hat{X}^n_k = X^n_k + \lambda\left(\frac{b}{n} + \frac{\sigma}{\sqrt{n}}\epsilon_k\right) \) for some \( \lambda \in [0, 1]\). Taking expectation under \( E_n \) on both sides and using the above inequalities, we obtain

\[
E_n|X^n_{k+1}|^{2m} \leq E_n|X^n_k|^{2m} + \frac{2mK'}{n}E_n(|X^n_k|^{2m-2} + |X^n_k|^{2m}) + \frac{2m(2m - 1)K'}{n}E_n(|X^n_k|^{2m-2} + |X^n_k|^{2m})
\]

\[
= E_n|X^n_k|^{2m} + \frac{4m^2K'}{n}E_n(|X^n_k|^{2m-2} + |X^n_k|^{2m}).
\]

Noticing further that \( x^{2m-2} \leq 1 + x^{2m} \), we obtain

\[
E_n|X^n_{k+1}|^{2m} \leq E_n|X^n_k|^{2m} + \frac{4m^2K'}{n}E_n[1 + 2|X^n_k|^{2m}]
\]

Hence, we can find a constant \( C > 0 \), depending only upon \( m \) and \( K' \), such that

\[
E_n|X^n_{k+1}|^{2m} \leq \frac{C}{n} + \left(1 + \frac{C}{n}\right)E_n|X^n_k|^{2m}.
\]
This implies that
\[ E_n |X^n_t|^2 \leq (1 + \frac{C}{n})^k (1 + E_n |X^n_0|^2)^{2m}. \]

Thus \( A = \sup_n (1 + \frac{C}{n})^n \) is our choice. \( \square \)

**Proof of Proposition 1.** First, we define
\[ V(x, t, \beta) = E[g(X(1), \beta) | X(t) = x], \]
then \( V \) satisfies the following partial differential equation:
\[ \frac{1}{2} \sigma(x, \beta)^2 V_{zz}(x, t, \beta) + b(x, \beta) V_z(x, t, \beta) + V_t(x, t, \beta) = 0, \quad (18) \]
\[ V(x, 1, \beta) = g(x, \beta). \quad (19) \]

We argue that \( V \) is continuously differential up to the third order and all these derivatives satisfy a polynomial growth condition. Following the argument used in Theorem 5.5 of Friedman (1975, pp. 122) and Theorem 1 of Gihman and Skorohod (1972, pp. 61), we conclude that \( V \) is continuously differentiable with respect to \( x \) up to the sixth order and that all of these derivatives satisfy a polynomial growth condition. Moreover, since \( V \) satisfies the above the PDE, \( V_t, V_{tt} \) and \( V_{ttt} \) can be expressed as functions of the partial derivatives of \( V \) with respect to \( x \) up to at most the sixth order. We deduce that \( V \) is also continuously differentiable with respect to \( t \) up to the third order. This plus the fact that \( b, \sigma \), and all of their derivatives satisfy a polynomial growth condition leads us to further deduce that all of the derivatives with respect to \( x \) and \( t \) up to the third order satisfy a polynomial growth condition.

Next, we define
\[ V^n(X^n_k, \frac{k}{n}, \beta) = E_n[g(X^n_n, \beta) | X^n_k], \]
then \( V^n \) satisfies the following recurrent equation:
\[ V^n(X^n_k, \frac{k}{n}, \beta) = \frac{1}{2} V^n(X^n_{k+1}, \frac{k+1}{n}, \beta) + \frac{1}{2} V^n(X^n_{k-1}, \frac{k+1}{n}, \beta), \quad (20) \]
where
\[ X^n_{k+1} = X^n_k + \frac{b(X^n_k, \beta) + \sigma(X^n_k, \beta)}{\sqrt{n}}, \]
\[ X^n_{k-1} = X^n_k + \frac{b(X^n_k, \beta) - \sigma(X^n_k, \beta)}{\sqrt{n}}. \]

Letting \( e^n_k = V(X^n_k, \frac{k}{n}, \beta) - V^n(X^n_k, \frac{k}{n}, \beta) \), we need to show that \( |e^n_k| \leq \frac{C}{n} (1 + |X_0|^2) \) for some constants \( C \) and \( \gamma' \). This can be done as follows.
Applying the Taylor expansion to $V$ and use the fact that $V$ satisfies (18), we get
\[
\frac{1}{2} V(X^+_{k+1}, \frac{k+1}{n}, \beta) + \frac{1}{2} V(X^-_{k+1}, \frac{k+1}{n}, \beta) = V(X^+_k, \frac{k}{n}, \beta) - \frac{1}{n^2} h(X^+_k) - \gamma^n_k,
\]  
(21)

where $h = -\frac{1}{2} V_{xx} \sigma^2 - V_{xt} \delta - \frac{1}{2} V_{tt}$ and $\gamma^n_k$ is the residual term in the order of $n^{-2}$ or higher. Specifically, define
\[
f_{k,n}(s) = \frac{1}{2} V(X^+_k + s(X^+_{k+1} - X^+_k), \frac{k+s}{n}, \beta) + \frac{1}{2} V(X^-_k + s(X^-_{k+1} - X^-_k), \frac{k+s}{n}, \beta),
\]
then
\[
\gamma^n_k = -\frac{1}{6} \int_0^1 (1-s)^3 f^{(3)}_{k,n}(s) ds.
\]

where $f^{(3)}$ denotes the third order derivative of $f$. Equation (21) allows us to derive an recurrent equation for $\epsilon^n_k$,
\[
\epsilon^n_k = \epsilon^n_{k+1} + \epsilon^n_{k-1}) + \frac{1}{n^2} h(X^+_k, \frac{k}{n}, \beta) + \gamma^n_k.
\]  
(22)

By direct substitution, we can represent $\epsilon^n_0$ in terms of the following expectation
\[
\epsilon^n_0 = E_n \left[ \frac{1}{n^2} \sum_{m=0}^{n-1} h(X^+_m, \frac{m}{n}, \beta) + \sum_{m=0}^{n} \gamma^n_m \right].
\]

Since $h$ satisfies a polynomial growth condition, there exists an integer $q > 0$ and a constant $K'$, such that
\[
|h(X^+_m, \frac{m}{n}, \beta)| \leq K'(1 + |X^+_m|^{2q}).
\]

Applying Lemma 1, we have
\[
E_n |h(X^+_m, \frac{m}{n}, \beta)| \leq K'(1 + A(1 + E[|X^+_m|^{2q}]), \quad 0 \leq m \leq n.
\]

Thus, there exist constants $C$ and $\gamma_1 > 0$, independent of $n$ and $\beta$, such that
\[
E_n \frac{1}{n^2} \sum_{m=0}^{n-1} h(X^+_m, \frac{m}{n}, \beta) \leq \frac{C_1}{n} (1 + |X^+_0|^\gamma_1).
\]

Similar treatment can be applied to $\gamma^n_k$, since it is in the order of $n^{-2}$ or higher. Hence, we obtain
\[
|\epsilon^n_0| \leq \frac{C}{n} (1 + |X^+_0|^{\gamma'})
\]
for some constant $C$ and $\gamma'$. The proof for the second part is exactly the same, if we treat $(X, D)$ as a two-dimensional diffusion process, where $D = \frac{\partial X}{\partial \beta}$. More specifically, denote $\frac{\partial F}{\partial \beta}$ by $H$, we have
\[
H(\beta) = E[\gamma(X(1), D(1), \beta)|X(0) = x, D(0) = 0],
\]
25
where \( \tilde{g}(x, y, \beta) = g_z(x, \beta)y + g_\beta(x, \beta). \)

**Proof of Proposition 2.** Without loss of generality, we assume that

\[
 f_t(X, \beta) = g_t(X, \beta) - E^\beta[g_t(X, \beta)|X_{t-1}],
\]

where \( g_t(X, \beta) = g(X_t, X_{t-1}, \beta) \) satisfying the sufficient conditions of Proposition 1. In particular, there exist constants \( K, \gamma > 0 \) such that

\[
 |g_t(X, \beta)| \leq K(1 + |X_{t-1}|^\gamma + |X_t|^{\gamma}).
\]

Following Lemma 1 and Proposition 1, we can find constants \( C', \gamma' > 0 \), independent of \( n \) and \( \beta \), such that

\[
 |f_t(X, \beta) - f_t^n(X, \beta)| \leq \frac{C'}{n}(1 + |X_{t-1}|^{\gamma'}),
\]

\[
 |E^\beta[g_t(X, \beta)|X_{t-1}]| \leq C'(1 + |X_{t-1}|^{\gamma'}),
\]

\[
 |E^\beta[g(X^n_{t-1}, X_{t-1}, \beta)|X_{t-1}]| \leq C'(1 + |X_{t-1}|^{\gamma'}),
\]

where \( X^n \) is the approximation for \( X \) between \( t-1 \) and \( t \) with \( X^n_{t-1} = X_{t-1} \). The above inequalities imply that

\[
 |R(\beta) - R^n(\beta)| \leq \frac{C'}{n}(1 + E|X_{t-1}|^{\gamma'}),
\]

\[
 |f_t(X, \beta)| \leq K'(1 + |X_t|^{\gamma} + |X_{t-1}|^{\gamma}),
\]

\[
 |f_t^n(X, \beta)| \leq K'(1 + |X_t|^{\gamma} + |X_{t-1}|^{\gamma}),
\]

for some constants \( K', \gamma' > 0 \). Denoting \( f_t(X, \beta) \) by \( f_t^\beta \) and \( f_t^n(X, \beta) \) by \( f_t^n\beta \), we consider the following,

\[
 |H(\beta) - H^n(\beta)| = |(E f_t^\beta)^\top W_0 (E f_t^\beta) - (E f_t^n\beta)^\top W^n (E f_t^n\beta)|
\]

\[
 \leq E|f_t^\beta - f_t^n\beta| \quad || W_0 || E|f_t^\beta| + E|f_t^n\beta| \quad || W_0 - W^n || E|f_t^\beta|
\]

\[
 + E|f_t^n\beta| \quad || W^n || E|f_t^n\beta - f_t^\beta|
\]

\[
 \leq \frac{C'K'}{n} E(1 + |X_{t-1}|^{\gamma'}) E(1 + |X_{t-1}|^{\gamma} + |X_t|^{\gamma}) || W_0 ||
\]

\[
 + K'^2 E(1 + |X_{t-1}|^{\gamma} + |X_t|^{\gamma})^2 || W_0 - W^n ||
\]

\[
 + \frac{C'K'}{n} E(1 + |X_{t-1}|^{\gamma'}) E(1 + |X_{t-1}|^{\gamma} + |X_t|^{\gamma}) || W^n || .
\]

Since \( E|X_{t-1}|^{\gamma} = E|X_t|^{\gamma} = E|Y|^{\gamma} < \infty \), and \( E|X_{t-1}|^{\gamma'} = E|Y|^{\gamma'} < \infty \), \( R^n \) and \( H^n \) converge to \( R \) and \( H \) uniformly on \( \Theta \).

Similar argument can be used to prove that \( J^n \) converges to \( J \) uniformly on \( \Theta \).
Table 1: (a) Finite sample properties of $\hat{\beta}_T$ for $T = 520$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std. dev.)</th>
<th>Var($\hat{\beta}_T$) (Std. dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std. dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 0.80</td>
<td>1.301 (0.623)</td>
<td>0.215 (0.464)</td>
<td>0.2806 (0.132)</td>
<td>0.124 (25.62)</td>
<td>0.212 (16.62)</td>
<td>0.268 (12.52)</td>
<td>217.2 (0.00)</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.1001 (0.311e-2)</td>
<td>0.954e-5 (0.309e-2)</td>
<td>0.989e-5 (0.119e-5)</td>
<td>0.008 (-0.449)</td>
<td>0.054 (0.410)</td>
<td>0.106 (0.447)</td>
<td>22.24 (0.27)</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.0994 (0.126e-1)</td>
<td>0.162e-3 (0.127e-1)</td>
<td>0.201e-3 (0.101e-2)</td>
<td>0.014 (-1.641)</td>
<td>0.034 (0.068)</td>
<td>0.068 (0.84)</td>
<td>25.84 (0.15)</td>
</tr>
</tbody>
</table>

Table 1: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 780$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std. dev.)</th>
<th>Var($\hat{\beta}_T$) (Std. dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std. dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 0.80</td>
<td>1.102 (0.428)</td>
<td>0.143 (0.379)</td>
<td>0.187 (0.384e-1)</td>
<td>0.080 (15.73)</td>
<td>0.150 (10.26)</td>
<td>0.200 (7.453)</td>
<td>135.8 (0.00)</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.1000 (0.241e-2)</td>
<td>0.636e-5 (0.252e-2)</td>
<td>0.659e-5 (0.797e-6)</td>
<td>0.000 (-2.247)</td>
<td>0.038 (-1.231)</td>
<td>0.106 (0.447)</td>
<td>18.56 (0.49)</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.1003 (0.107e-1)</td>
<td>0.168e-3 (0.104e-1)</td>
<td>0.133e-3 (0.671e-3)</td>
<td>0.016 (-1.348)</td>
<td>0.048 (-0.205)</td>
<td>0.106 (0.447)</td>
<td>14.24 (0.78)</td>
</tr>
</tbody>
</table>

Table 1: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 1040$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std. dev.)</th>
<th>Var($\hat{\beta}_T$) (Std. dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std. dev.)</th>
<th>0.01 tail$^a$ (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 0.80</td>
<td>0.993 (0.343)</td>
<td>0.107 (0.328)</td>
<td>0.108 (0.377e-1)</td>
<td>0.034 (5.393)</td>
<td>0.102 (5.334)</td>
<td>0.150 (3.727)</td>
<td>92.80 (0.00)</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.0999 (0.215e-2)</td>
<td>0.477e-5 (0.218e-2)</td>
<td>0.492e-5 (0.423e-6)</td>
<td>0.010 (0.000)</td>
<td>0.044 (-0.615)</td>
<td>0.092 (-0.596)</td>
<td>20.80 (0.36)</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.1003 (0.951e-2)</td>
<td>0.809e-4 (0.899e-2)</td>
<td>0.753e-4 (0.737e-4)</td>
<td>0.014 (1.846)</td>
<td>0.068 (1.192)</td>
<td>0.116 (0.89)</td>
<td>11.84 (0.89)</td>
</tr>
</tbody>
</table>

$^a$ Asymptotic t-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{\beta} - p)/\sqrt{p(1-p)/N}^{1/2}$, where $N = 500$ is the number of replications and $\hat{\beta}$ is the estimated proportion.
Table 2: (a) Finite sample properties of $\hat{\beta}_T$ for $T = 260$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail* (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.919</td>
<td>1.994</td>
<td>2.185</td>
<td>0.06</td>
<td>0.122</td>
<td>0.178</td>
<td>152.5</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.1003</td>
<td>0.197e-4</td>
<td>0.200e-4</td>
<td>0.00</td>
<td>0.06</td>
<td>0.116</td>
<td>20.16</td>
</tr>
<tr>
<td></td>
<td>(0.455e-2)</td>
<td>(0.443e-2)</td>
<td>(0.417e-5)</td>
<td>(-2.247)</td>
<td>(1.026)</td>
<td>(1.192)</td>
<td>(0.39)</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000</td>
<td>0.131e-4</td>
<td>0.113e-4</td>
<td>0.008</td>
<td>0.044</td>
<td>0.092</td>
<td>15.20</td>
</tr>
<tr>
<td></td>
<td>(0.352e-2)</td>
<td>(0.363e-2)</td>
<td>(0.803e-5)</td>
<td>(-0.899)</td>
<td>(-0.615)</td>
<td>(-0.596)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 520$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail* (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.274</td>
<td>0.997</td>
<td>0.993</td>
<td>0.024</td>
<td>0.068</td>
<td>0.132</td>
<td>36.21</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.0999</td>
<td>0.984e-5</td>
<td>0.104e-5</td>
<td>0.014</td>
<td>0.054</td>
<td>0.098</td>
<td>15.28</td>
</tr>
<tr>
<td></td>
<td>(0.326e-2)</td>
<td>(0.313e-2)</td>
<td>(0.146e-5)</td>
<td>(0.899)</td>
<td>(0.410)</td>
<td>(-0.149)</td>
<td>(0.71)</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000</td>
<td>0.658e-5</td>
<td>0.589e-5</td>
<td>0.004</td>
<td>0.032</td>
<td>0.078</td>
<td>22.16</td>
</tr>
<tr>
<td></td>
<td>(0.233e-2)</td>
<td>(0.256e-2)</td>
<td>(0.295e-5)</td>
<td>(-1.348)</td>
<td>(-1.846)</td>
<td>(-1.639)</td>
<td>(0.27)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 780$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.error)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\overline{\text{Var}}(\hat{\beta}_T$) (Std.error)</th>
<th>0.01 tail* (t-stat.)</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>4.00</td>
<td>4.124</td>
<td>0.665</td>
<td>0.636</td>
<td>0.02</td>
<td>0.046</td>
<td>0.114</td>
<td>21.92</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.0999</td>
<td>0.656e-5</td>
<td>0.693e-5</td>
<td>0.008</td>
<td>0.042</td>
<td>0.100</td>
<td>9.28</td>
</tr>
<tr>
<td></td>
<td>(0.258e-2)</td>
<td>(0.256e-2)</td>
<td>(0.765e-6)</td>
<td>(-0.449)</td>
<td>(-0.821)</td>
<td>(0.000)</td>
<td>(0.97)</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.10</td>
<td>0.1000</td>
<td>0.439e-5</td>
<td>0.397e-5</td>
<td>0.004</td>
<td>0.050</td>
<td>0.102</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>(0.204e-2)</td>
<td>(0.209e-2)</td>
<td>(0.156e-7)</td>
<td>(-1.348)</td>
<td>(0.000)</td>
<td>(0.152)</td>
<td>(0.87)</td>
<td></td>
</tr>
</tbody>
</table>

* Asymptotic t-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$, where $N = 500$ is the number of replications and $\hat{p}$ is the estimated proportion.
Table 3: (a) Finite sample properties of $\hat{\beta}_T$ for $T = 260$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail a</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 40.0</td>
<td>41.24 (6.104)</td>
<td>40.86 (6.592)</td>
<td>41.29 (13.692)</td>
<td>0.02</td>
<td>0.058</td>
<td>0.118</td>
<td>22.72</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.1004 (0.636e-2)</td>
<td>0.374e-4 (0.611e-2)</td>
<td>0.408e-4 (0.109e-4)</td>
<td>0.028</td>
<td>0.062</td>
<td>0.104</td>
<td>20.32</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.1000 (0.354e-3)</td>
<td>0.128e-6 (0.358e-3)</td>
<td>0.131e-6 (0.284e-7)</td>
<td>0.004</td>
<td>0.048</td>
<td>0.098</td>
<td>25.52</td>
</tr>
</tbody>
</table>

Table 3: (b) Finite sample properties of $\hat{\beta}_T$ for $T = 520$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail a</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 40.0</td>
<td>40.47 (4.427)</td>
<td>20.42 (4.519)</td>
<td>19.75 (4.651)</td>
<td>0.012</td>
<td>0.046</td>
<td>0.088</td>
<td>16.16</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.1004 (0.447e-2)</td>
<td>0.187e-4 (0.432e-2)</td>
<td>0.200e-4 (0.371e-5)</td>
<td>0.022</td>
<td>0.058</td>
<td>0.098</td>
<td>10.40</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.1000 (0.253e-3)</td>
<td>0.641e-7 (0.253e-3)</td>
<td>0.661e-7 (0.104e-7)</td>
<td>0.014</td>
<td>0.058</td>
<td>0.106</td>
<td>20.64</td>
</tr>
</tbody>
</table>

Table 3: (c) Finite sample properties of $\hat{\beta}_T$ for $T = 780$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{\beta}_T$ (Std.dev.)</th>
<th>Var($\hat{\beta}_T$) (Std.dev.)</th>
<th>$\text{Var}(\hat{\beta}_T)$ (Std.dev.)</th>
<th>0.01 tail a</th>
<th>0.05 tail (t-stat.)</th>
<th>0.10 tail (t-stat.)</th>
<th>$\chi^2(19)$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$ 40.0</td>
<td>40.17 (3.699)</td>
<td>13.62 (3.691)</td>
<td>13.34 (2.676)</td>
<td>0.008</td>
<td>0.046</td>
<td>0.106</td>
<td>11.84</td>
</tr>
<tr>
<td>$\sigma$ 0.10</td>
<td>0.1007 (0.380e-2)</td>
<td>0.125e-4 (0.353e-2)</td>
<td>0.136e-4 (0.216e-5)</td>
<td>0.012</td>
<td>0.060</td>
<td>0.112</td>
<td>14.88</td>
</tr>
<tr>
<td>$\theta$ 0.10</td>
<td>0.0999 (0.202e-3)</td>
<td>0.427e-7 (0.207e-3)</td>
<td>0.434e-7 (0.056e-7)</td>
<td>0.004</td>
<td>0.048</td>
<td>0.098</td>
<td>13.52</td>
</tr>
</tbody>
</table>

a Asymptotic t-statistics for the hypothesis that the true proportion is $p$ is calculated as $(\hat{p} - p) / [p(1-p)/N]^{1/2}$, where $N = 500$ is the number of replications and $\hat{p}$ is the estimated proportion.
Table 4: (a) First moment approximations $F(\beta) = E^\beta[X_{t+1}|X_t]$.  
\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P$</th>
<th>$\frac{\partial P}{\partial \kappa}$</th>
<th>$\frac{\partial P}{\partial \sigma}$</th>
<th>$\frac{\partial P}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.8489e-1</td>
<td>0.6075e-2</td>
<td>0.000</td>
<td>0.1528e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.8489e-1</td>
<td>0.6069e-2</td>
<td>0.000</td>
<td>0.1528e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.8499e-1</td>
<td>0.6065e-2</td>
<td>0.000</td>
<td>0.1527e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.8488e-1</td>
<td>0.6062e-2</td>
<td>0.000</td>
<td>0.1527e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.8488e-1</td>
<td>0.6061e-2</td>
<td>0.000</td>
<td>0.1527e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.8488e-1</td>
<td>0.6059e-2</td>
<td>0.000</td>
<td>0.1526e-1</td>
</tr>
</tbody>
</table>

Table 4: (b) Second moment approximations $F(\beta) = E^\beta[X_{t+1}^2|X_t]$.  
\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P$</th>
<th>$\frac{\partial P}{\partial \kappa}$</th>
<th>$\frac{\partial P}{\partial \sigma}$</th>
<th>$\frac{\partial P}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1559e-2</td>
<td>0.2626e-2</td>
</tr>
<tr>
<td>10</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1559e-2</td>
<td>0.2626e-2</td>
</tr>
<tr>
<td>12</td>
<td>0.7596e-2</td>
<td>0.1037e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>14</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>16</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1560e-2</td>
<td>0.2627e-2</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7596e-2</td>
<td>0.1036e-2</td>
<td>0.1561e-2</td>
<td>0.2628e-2</td>
</tr>
</tbody>
</table>

Table 4: (c) Third moment approximations $F(\beta) = E^\beta[X_{t+1}^3|X_t]$.  
\[ \kappa = 0.8, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P$</th>
<th>$\frac{\partial P}{\partial \kappa}$</th>
<th>$\frac{\partial P}{\partial \sigma}$</th>
<th>$\frac{\partial P}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7133e-3</td>
<td>0.1340e-3</td>
<td>0.4155e-3</td>
<td>0.3564e-3</td>
</tr>
<tr>
<td>10</td>
<td>0.7134e-3</td>
<td>0.1340e-3</td>
<td>0.4163e-3</td>
<td>0.3566e-3</td>
</tr>
<tr>
<td>12</td>
<td>0.7135e-3</td>
<td>0.1399e-3</td>
<td>0.4169e-3</td>
<td>0.3567e-3</td>
</tr>
<tr>
<td>14</td>
<td>0.7136e-3</td>
<td>0.1399e-3</td>
<td>0.4174e-3</td>
<td>0.3569e-3</td>
</tr>
<tr>
<td>16</td>
<td>0.7136e-3</td>
<td>0.1399e-3</td>
<td>0.4180e-3</td>
<td>0.3571e-3</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7138e-3</td>
<td>0.1399e-3</td>
<td>0.4187e-3</td>
<td>0.3574e-3</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 5: (a) First moment approximations $F(\beta) = E^\beta[X_{t+1}|X_t]$. 

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$F_{\beta x}$</th>
<th>$F_{\beta \sigma}$</th>
<th>$F_{\beta \sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.1487e-2</td>
<td>0.3594e-3</td>
<td>0.000</td>
<td>0.7436e-1</td>
</tr>
<tr>
<td>10</td>
<td>8.1486e-2</td>
<td>0.3588e-3</td>
<td>0.000</td>
<td>0.7427e-1</td>
</tr>
<tr>
<td>12</td>
<td>8.1485e-2</td>
<td>0.3582e-3</td>
<td>0.000</td>
<td>0.7421e-1</td>
</tr>
<tr>
<td>14</td>
<td>8.1485e-2</td>
<td>0.3577e-3</td>
<td>0.000</td>
<td>0.7416e-1</td>
</tr>
<tr>
<td>16</td>
<td>8.1484e-2</td>
<td>0.3569e-3</td>
<td>0.000</td>
<td>0.7411e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>8.1481e-2</td>
<td>0.3561e-3</td>
<td>0.000</td>
<td>0.7403e-1</td>
</tr>
</tbody>
</table>

Table 5: (b) Second moment approximations $F(\beta) = E^\beta[X_{t+1}^2|X_t]$. 

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$F_{\beta x}$</th>
<th>$F_{\beta \sigma}$</th>
<th>$F_{\beta \sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.6654e-2</td>
<td>0.5834e-4</td>
<td>0.2902e-3</td>
<td>0.1213e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.6654e-2</td>
<td>0.5823e-4</td>
<td>0.2898e-3</td>
<td>0.1213e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.6654e-3</td>
<td>0.5810e-4</td>
<td>0.2895e-3</td>
<td>0.1211e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.6653e-3</td>
<td>0.5802e-4</td>
<td>0.2892e-3</td>
<td>0.1210e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.6653e-3</td>
<td>0.5796e-4</td>
<td>0.2891e-3</td>
<td>0.1209e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.6653e-3</td>
<td>0.5780e-4</td>
<td>0.2879e-3</td>
<td>0.1207e-1</td>
</tr>
</tbody>
</table>

Table 5: (c) Third moment approximations $F(\beta) = E^\beta[X_{t+1}^3|X_t]$. 

\[ \kappa = 4, \sigma = 0.1, \theta = 0.1 \text{ and } X_t = 0.08. \]

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$F_{\beta x}$</th>
<th>$F_{\beta \sigma}$</th>
<th>$F_{\beta \sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.5446e-3</td>
<td>0.7125e-5</td>
<td>0.7110e-4</td>
<td>0.1487e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.5446e-3</td>
<td>0.7110e-5</td>
<td>0.7097e-4</td>
<td>0.1485e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.5446e-3</td>
<td>0.7100e-5</td>
<td>0.7087e-4</td>
<td>0.1484e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.5445e-3</td>
<td>0.7093e-5</td>
<td>0.7079e-4</td>
<td>0.1484e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.5445e-3</td>
<td>0.7088e-5</td>
<td>0.7072e-4</td>
<td>0.1483e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.5445e-3</td>
<td>0.7068e-5</td>
<td>0.7054e-4</td>
<td>0.1481e-1</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 6: (a) First moment approximations $F(\beta) = E^\beta[X_{t+1}|X_t]$.

$\kappa = 40$, $\sigma = 0.1$, $\theta = 0.1$ and $X_t = 0.08$.

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.9109e-1</td>
<td>0.1845e-3</td>
<td>0.000</td>
<td>0.5516</td>
</tr>
<tr>
<td>10</td>
<td>0.9096e-1</td>
<td>0.1812e-3</td>
<td>0.000</td>
<td>0.5448</td>
</tr>
<tr>
<td>12</td>
<td>0.9085e-1</td>
<td>0.1795e-3</td>
<td>0.000</td>
<td>0.5401</td>
</tr>
<tr>
<td>14</td>
<td>0.9079e-1</td>
<td>0.1791e-3</td>
<td>0.000</td>
<td>0.5387</td>
</tr>
<tr>
<td>16</td>
<td>0.9075e-1</td>
<td>0.1788e-3</td>
<td>0.000</td>
<td>0.5373</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.9073e-1</td>
<td>0.1782e-3</td>
<td>0.000</td>
<td>0.5366</td>
</tr>
</tbody>
</table>

Table 6: (b) Second moment approximations $F(\beta) = E^\beta[X_{t+1}^2|X_t]$.

$\kappa = 40$, $\sigma = 0.1$, $\theta = 0.1$ and $X_t = 0.08$.

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.8301e-2</td>
<td>0.3415e-4</td>
<td>0.1819e-3</td>
<td>0.1010</td>
</tr>
<tr>
<td>10</td>
<td>0.8283e-2</td>
<td>0.3356e-4</td>
<td>0.1789e-3</td>
<td>0.0998</td>
</tr>
<tr>
<td>12</td>
<td>0.8264e-2</td>
<td>0.3304e-4</td>
<td>0.1771e-3</td>
<td>0.0990</td>
</tr>
<tr>
<td>14</td>
<td>0.8256e-2</td>
<td>0.3279e-4</td>
<td>0.1754e-3</td>
<td>0.0983</td>
</tr>
<tr>
<td>16</td>
<td>0.8251e-2</td>
<td>0.3231e-4</td>
<td>0.1741e-3</td>
<td>0.0979</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.8241e-2</td>
<td>0.3223e-4</td>
<td>0.1714e-3</td>
<td>0.0974</td>
</tr>
</tbody>
</table>

Table 6: (c) Third moment approximations $F(\beta) = E^\beta[X_{t+1}^3|X_t]$.

$\kappa = 40$, $\sigma = 0.1$, $\theta = 0.1$ and $X_t = 0.08$.

<table>
<thead>
<tr>
<th>n</th>
<th>$F$</th>
<th>$\frac{\partial F}{\partial \kappa}$</th>
<th>$\frac{\partial F}{\partial \sigma}$</th>
<th>$\frac{\partial F}{\partial \theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.7582e-3</td>
<td>0.4621e-5</td>
<td>0.4962e-4</td>
<td>0.1383e-1</td>
</tr>
<tr>
<td>10</td>
<td>0.7557e-3</td>
<td>0.4546e-5</td>
<td>0.4825e-4</td>
<td>0.1371e-1</td>
</tr>
<tr>
<td>12</td>
<td>0.7542e-3</td>
<td>0.4489e-5</td>
<td>0.4753e-4</td>
<td>0.1363e-1</td>
</tr>
<tr>
<td>14</td>
<td>0.7523e-3</td>
<td>0.4452e-5</td>
<td>0.4695e-4</td>
<td>0.1351e-1</td>
</tr>
<tr>
<td>16</td>
<td>0.7511e-3</td>
<td>0.4429e-5</td>
<td>0.4684e-4</td>
<td>0.1344e-1</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.7493e-3</td>
<td>0.4406e-5</td>
<td>0.4673e-4</td>
<td>0.1331e-1</td>
</tr>
</tbody>
</table>

Note: $n = \infty$ corresponds to the values and derivatives of the actual conditional moments.
Table 7: (a) Approximations for the objective function.

\( T = 520, \kappa = 0.8, \sigma = 0.1, \text{ and } \theta = 0.1. \)

<table>
<thead>
<tr>
<th>n</th>
<th>OBJ</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.260921</td>
<td>0.33350e-1</td>
<td>-30.6410</td>
<td>-1.8424</td>
</tr>
<tr>
<td>10</td>
<td>0.260920</td>
<td>0.33347e-1</td>
<td>-30.6410</td>
<td>-1.8420</td>
</tr>
<tr>
<td>12</td>
<td>0.260920</td>
<td>0.33345e-1</td>
<td>-30.5410</td>
<td>-1.8417</td>
</tr>
<tr>
<td>14</td>
<td>0.260919</td>
<td>0.33343e-1</td>
<td>-30.6410</td>
<td>-1.8413</td>
</tr>
<tr>
<td>16</td>
<td>0.260919</td>
<td>0.33341e-1</td>
<td>-30.6410</td>
<td>-1.8409</td>
</tr>
<tr>
<td>∞</td>
<td>0.260919</td>
<td>0.33338e-1</td>
<td>-30.6409</td>
<td>-1.8405</td>
</tr>
</tbody>
</table>

Table 7: (b) Approximations for the objective function.

\( T = 520, \kappa = 4, \sigma = 0.1, \text{ and } \theta = 0.1. \)

<table>
<thead>
<tr>
<th>n</th>
<th>OBJ</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.9002e-1</td>
<td>0.2137e-1</td>
<td>-15.3653</td>
<td>-4.7632</td>
</tr>
<tr>
<td>10</td>
<td>0.8998e-1</td>
<td>0.2133e-1</td>
<td>-15.3665</td>
<td>-4.7532</td>
</tr>
<tr>
<td>12</td>
<td>0.8994e-1</td>
<td>0.2130e-1</td>
<td>-15.3653</td>
<td>-4.7466</td>
</tr>
<tr>
<td>14</td>
<td>0.8992e-1</td>
<td>0.2128e-1</td>
<td>-15.3641</td>
<td>-4.7391</td>
</tr>
<tr>
<td>16</td>
<td>0.8991e-1</td>
<td>0.2126e-1</td>
<td>-15.3631</td>
<td>-4.7311</td>
</tr>
<tr>
<td>∞</td>
<td>0.8987e-1</td>
<td>0.2118e-1</td>
<td>-15.3607</td>
<td>-4.7298</td>
</tr>
</tbody>
</table>

Table 7: (c) Approximations for the objective function.

\( T = 520, \kappa = 40, \sigma = 0.1, \text{ and } \theta = 0.1. \)

<table>
<thead>
<tr>
<th>n</th>
<th>OBJ</th>
<th>GRD(1)</th>
<th>GRD(2)</th>
<th>GRD(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.2024</td>
<td>0.2108e-1</td>
<td>-25.886</td>
<td>-100.351</td>
</tr>
<tr>
<td>10</td>
<td>0.2016</td>
<td>0.2088e-1</td>
<td>-25.821</td>
<td>-98.891</td>
</tr>
<tr>
<td>12</td>
<td>0.2011</td>
<td>0.2076e-1</td>
<td>-25.781</td>
<td>-97.746</td>
</tr>
<tr>
<td>14</td>
<td>0.2007</td>
<td>0.2063e-1</td>
<td>-25.756</td>
<td>-96.871</td>
</tr>
<tr>
<td>16</td>
<td>0.2005</td>
<td>0.2054e-1</td>
<td>-25.736</td>
<td>-95.981</td>
</tr>
<tr>
<td>∞</td>
<td>0.1991</td>
<td>0.2019e-1</td>
<td>-25.613</td>
<td>-94.541</td>
</tr>
</tbody>
</table>

Note: \( n = \infty \) corresponds to the values and derivatives of the actual objective functions.
Table 8: Comparisons between $\hat{\beta}_T$ and $\hat{\beta}_T^n$.

$\kappa = 4$, $\sigma = 0.1$, $\theta = 0.1$, $T = 520$, and $n = 12$

<table>
<thead>
<tr>
<th>$\bar{\kappa}_T$</th>
<th>$\bar{\kappa}_T^n$</th>
<th>$\bar{\beta}_T$</th>
<th>$\bar{\beta}_T^n$</th>
<th>$\hat{\beta}_T$</th>
<th>$\hat{\beta}_T^n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6514925 (1.0211)</td>
<td>4.6514925 (1.0237)</td>
<td>0.1013149 (0.3475e-2)</td>
<td>0.101348 (0.3480e-2)</td>
<td>0.1011897 (0.2212e-2)</td>
<td>0.1011866 (0.2198e-2)</td>
</tr>
<tr>
<td>4.1857486 (1.0514)</td>
<td>4.1857485 (1.0523)</td>
<td>0.0990128 (0.3075e-2)</td>
<td>0.0990119 (0.3077e-2)</td>
<td>0.1014964 (0.2388e-2)</td>
<td>0.1014965 (0.2375e-2)</td>
</tr>
<tr>
<td>4.1202013 (0.9060)</td>
<td>4.1202037 (0.9064)</td>
<td>0.1024504 (0.3053e-2)</td>
<td>0.1024501 (0.3056e-2)</td>
<td>0.1014151 (0.2426e-2)</td>
<td>0.1014173 (0.2414e-2)</td>
</tr>
<tr>
<td>4.4843505 (1.0332)</td>
<td>4.4843505 (1.0339)</td>
<td>0.1035081 (0.3268e-2)</td>
<td>0.1035085 (0.3271e-2)</td>
<td>0.0970444 (0.2280e-2)</td>
<td>0.0970441 (0.2267e-2)</td>
</tr>
<tr>
<td>3.9555679 (1.0234)</td>
<td>3.9555697 (1.0243)</td>
<td>0.0968098 (0.3375e-2)</td>
<td>0.0968098 (0.3379e-2)</td>
<td>0.0977091 (0.2414e-2)</td>
<td>0.0977099 (0.2402e-2)</td>
</tr>
<tr>
<td>3.8589336 (0.9804)</td>
<td>3.8589386 (0.9811)</td>
<td>0.1022397 (0.3325e-2)</td>
<td>0.1022398 (0.3327e-2)</td>
<td>0.1013215 (0.2702e-2)</td>
<td>0.1013174 (0.2689e-2)</td>
</tr>
<tr>
<td>3.9622223 (1.0068)</td>
<td>3.9622223 (1.0072)</td>
<td>0.1116794 (0.3679e-2)</td>
<td>0.1116777 (0.3682e-2)</td>
<td>0.0969922 (0.2773e-2)</td>
<td>0.0969948 (0.2758e-2)</td>
</tr>
<tr>
<td>4.6551534 (1.0248)</td>
<td>4.6551534 (1.0266)</td>
<td>0.1015327 (0.3441e-2)</td>
<td>0.1015341 (0.3445e-2)</td>
<td>0.1026368 (0.2208e-2)</td>
<td>0.1026367 (0.2194e-2)</td>
</tr>
<tr>
<td>4.0139144 (0.9541)</td>
<td>4.0139144 (0.9549)</td>
<td>0.0987532 (0.3095e-2)</td>
<td>0.0987530 (0.3098e-2)</td>
<td>0.0984243 (0.2432e-2)</td>
<td>0.0984240 (0.2419e-2)</td>
</tr>
<tr>
<td>4.7025483 (0.9669)</td>
<td>4.7025483 (0.9686)</td>
<td>0.1081610 (0.3581e-2)</td>
<td>0.1081625 (0.3585e-2)</td>
<td>0.1014039 (0.2298e-2)</td>
<td>0.1014028 (0.2287e-2)</td>
</tr>
</tbody>
</table>

Note: The numbers in parentheses are the corresponding estimated standard errors.